[Counsel on signature page]

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

Abdi Nazemian, et al.,

Plaintiffs,

vs.

NVIDIA Corporation,

Defendant.

Andre Dubus III, et al.,

Plaintiffs,

vs.

NVIDIA Corporation,

Defendant.

Case No. 4:24-cv-01454-JST
Case No. 4:24-cv-02655-JST

BRIEF IN SUPPORT OF PLAINTIFFS' PROPOSED ESI AND PROTECTIVE ORDERS

I. Introduction

During the case management conference on August 29, 2024, the Court expressed a preference for the Northern District of California's Model Stipulated Orders regarding the discovery of electronically stored information ("ESI") in these cases. There are two such model orders: the "Model Patent Order" ("MPO"), tailored for patent cases, and the "Standard Model Order" ("SMO"), designed for standard, non-patent cases. Neither party proposes that the Court adopt one of the model orders outright. Instead, the *Nazemian* and *Dubus* Plaintiffs propose a common ESI protocol ("Plaintiffs' ESI Protocol") that is based on the SMO and includes additional provisions reflecting contemporary best practices in the ESI field. Conversely, NVIDIA's proposal is based on the MPO which, as explained below, is inappropriate for a non-patent case and was not designed for complex class cases such as these.

NVIDIA's proposed ESI protocol ("NVIDIA's ESI Protocol") is an ill-suited framework for this litigation. Unlike patent disputes, which are typically narrow and technical, these cases involve complex copyright claims concerning the misuse of Plaintiffs' creative works to train AI models. The MPO, created over a decade ago to address electronic discovery costs in patent cases, imposes rigid constraints that are inappropriate for cases of this breadth. In copyright litigation involving AI models, discovery will extend beyond a limited set of technical documents. These cases demand access to diverse ESI sources, such as datasets, internal communications, and records of corporate decisions regarding the misuse of Plaintiffs' copyrighted works. NVIDIA's ESI Protocol would impose strict limits—drawn from the MPO—on email searches, custodians, and search terms and would severely hinder Plaintiffs' ability to uncover relevant evidence. Such restrictions undermine the transparency, fairness, and proportionality required by the Federal Rules of Civil Procedure, particularly in cases of this importance and magnitude. *See Mironowski v. Ford Motor Co.*, No. 1:22-cv-00675-JLT-CDB, 2023 WL 2957858, at *2 (E.D. Cal. Apr. 14, 2023) (stating that the Northern District of California's preference for model orders to protect trade secrets and proprietary information applies only in patent cases).

NVIDIA's proposed Protective Order ("NVIDIA's Protective Order") is also inadequate for the needs of these cases, as it fails to distinguish between actual source code and other

materials requiring heightened protection. In contrast, Plaintiffs' proposed Protective Order ("Plaintiffs' Protective Order") clearly defines "Source Code" and includes additional provisions suited to these cases which are missing from NVIDIA's proposal.

II. An ESI Protocol Based on the MPO is Inappropriate in a Class Copyright Case

The MPO, created in 2011 to address rising electronic discovery costs in patent litigation, was designed for cases involving narrow technical issues such as prior art, inventorship, or patent prosecution history, where the relevant information is relatively limited. *See* Declaration of Randall R. Rader, dated September 19, 2024 ("Rader Decl."), ¶ 10; Declaration of Craig Ball, dated September 18, 2024 ("Ball Decl."), ¶ 10.[1] Courts often restrict discovery in patent cases based on the nature of the plaintiff's specific infringement contentions. *See, e.g.*, *Ceiva Logic, Inc. v. Amazon.com, Inc.*, No. 2:19-cv-09129-AB-MAA, 2021 WL 3284813, at *5 (C.D. Cal. June 16, 2021). Custodians in such cases are typically few, and email production is limited by the technical scope of the dispute. Such strict limitations would be unworkable in the class-action copyright cases here. *See* Rader Decl. ¶¶ 11-13; Ball Decl. ¶¶ 8-12. The MPO's custodian restrictions would hinder discovery of critical evidence across the organization, from engineers to executives, *see* Ball Decl. ¶ 10, and its limited search terms would make it nearly impossible to craft precise queries to uncover documents related to the misuse of copyrighted materials, *see* Ball Decl. ¶ 11.

Plaintiffs' concerns regarding stipulations based on the MPO are grounded in experience. In *Doe 1 v. GitHub, Inc.*, Case No. 4:22-cv-6823-JST (N.D. Cal.) ("*GitHub*"), this Court issued a similar discovery order limiting the scope of email ESI production to five search terms per custodian, across five custodians, while excluding non-email ESI from production. *See GitHub* ECF No. 259, at 1. Defendants in that case have exploited these limits: to date, over a year into litigation, one defendant has produced only 82 documents, and another 183—all mostly public or

[1] Judge Randall R. Rader is former Chief Judge of the United States Court of Appeals for the Federal Circuit who organized and participated in the E-Discovery Committee that drafted the Federal Circuit's Model Patent Order, on which the Northern District of California MPO is based. Rader Decl. ¶¶ 1, 3, 5; *id.* App. A. Craig Ball is an attorney, law professor, and ESI expert who has served in over 80 cases as court-appointed special master, neutral, consultant, or testifying expert in computer forensics and electronic discovery. Ball Dec. ¶ 1; *id.* App. A.

routine materials like organizational charts. *See id.* As a result, discovery has stalled, hindering Plaintiffs' ability to gather crucial evidence and proceed efficiently with the litigation. *See id.* Applying the same framework here would likely obstruct, rather than advance, the just, speedy, and inexpensive resolution of these cases. *See* Rader Decl. ¶ 12; Ball Decl. ¶ 6.

III. NVIDIA's ESI Protocol is Ill-Suited to This Litigation

NVIDIA's proposed ESI order is based on the MPO, as NVIDIA's counsel disclosed to Plaintiffs. But the MPO, or any version of it, is ill-suited for these class-action copyright cases. *See* Rader Decl. ¶¶ 11-13; *see also* Ball Decl. ¶¶ 12, 16 (noting the MPO's failure to address key issues like structured data, search methodology, and privilege logs). Unlike patent cases, these cases hinge on fact-intensive issues like fair use, necessitating broader discovery to evaluate factors including purpose, market impact, and scope of infringement. Yet, NVIDIA's ESI Protocol seeks to impose the strictest provisions of the MPO: (a) overly restrictive email discovery, including unreasonable custodial and search term limitations, (b) inadequate provisions for structured data, (c) an insufficient search methodology, (d) outdated mobile data discovery, and (e) a vague privilege log protocol. *See* NVIDIA's ESI Protocol §§ 6-8, 10-12, 18-19.

A. NVIDIA's Proposed Email Discovery Provisions Are Overly Restrictive

NVIDIA's proposal prohibits general email discovery requests and requires Plaintiffs to serve narrowly tailored email production requests for specific issues. NVIDIA ESI Protocol §§ 6-7. This proposal is not only unnecessarily restrictive, but it also runs counter to the principles of full and fair discovery, particularly in cases of this complexity. *See* Ball Decl. ¶ 9; Rader Decl. ¶ 13. Patent disputes typically involve a few discrete issues, such as questions of inventorship or the sharing of technical specifications among a limited number of individuals. The copyright infringement cases here involve a broader range of issues, including the unauthorized use of copyrighted materials on a massive scale, sophisticated datasets, and myriad decision-makers across multiple departments. Limiting email discovery would cripple Plaintiffs' ability to explore how NVIDIA made decisions regarding the training of its AI models, the extent to which NVIDIA executives and teams were aware of potential infringement, and whether efforts were made to mitigate such risks. Plaintiffs are entitled to uncover broad patterns of conduct, which cannot be

accomplished through narrowly tailored, issue-specific email requests. The very nature of this litigation, which spans multiple domains across legal, technical, and executive fields, requires an approach to email discovery that is more comprehensive than what NVIDIA's ESI Protocol provides.

NVIDIA's proposal also limits email discovery to eight custodians for all Plaintiffs across two distinct cases and five search terms per custodian (resulting in a mere 40 total search terms for all Plaintiffs). This draconian limitation is wholly inadequate for the needs of these cases. In litigation involving cutting-edge AI technology, relevant documents are not confined to a handful of individuals or departments. *See* Ball Decl. ¶ 10. The development and deployment of NVIDIA's models likely involved multiple teams, including research and development, data engineering, legal, and executive leadership. *See id.;* NVIDIA, Learning and Perception Research Group, https://research.nvidia.com/labs/lpr; Alex Melnichuk, Collaborative AI Development: Tips for Excellent Team Efficiency, N-iX, https://ncube.com/collaborative-ai-development-tips-for-excellent-team-efficiency.

Relevant documents are likely to be held by employees across these departments, each playing a different but crucial role in the decision-making process surrounding the use of Plaintiffs' copyrighted works. *See id.* By arbitrarily capping the number of custodians, NVIDIA's proposal would effectively exclude many key players from the discovery process, preventing Plaintiffs from tracing the full scope of internal discussions and decision-making that governed the use of these copyrighted works in NVIDIA's training datasets.

Moreover, limiting search terms to five per custodian ignores the complexity of modern litigation, where relevant documents often deploy varied terminology, industry-specific jargon, or acronyms that differ between teams or departments. In a case as multifaceted as this, multiple search terms are necessary to capture the full breadth of relevant documents. A five-term limit is particularly problematic in the context of AI development, where discussions may involve technical concepts, proprietary terms, or industry abbreviations. By artificially limiting search terms, NVIDIA's proposal would miss critical documents that use different phrasing, and this limitation would hamper Plaintiffs' ability to conduct thorough and effective discovery.

Plaintiffs' proposal, on the other hand, is far more reasonable and tailored to the needs of these cases. *See* Ball Decl. ¶¶ 14-15. Plaintiffs propose identifying 24 custodians, Plaintiffs ESI Order § 5.A, a figure a court in this District found to be appropriately aligned with the scope and complexity of a case involving massive amounts of data, systemic infringement, and multiple levels of decision-making across departments. *See In re OpenAI ChatGPT Litigation*, Lead Case No. 3:23-cv-3223-AMO (N.D. Cal.), ECF No. 175 (order in generative AI copyright infringement case allowing plaintiffs to seek discovery from 24 custodians). This approach recognizes that relevant documents are likely to be distributed across multiple departments and employees.

Similarly, Plaintiffs' proposal does not impose arbitrary limits on search terms. *See id.* § C. This flexibility is critical in complex cases where the precision and effectiveness of discovery often hinge on the ability to use a variety of search terms that can capture all relevant documents. Plaintiffs do not suggest a discovery free-for-all, but rather a proportional and measured approach that allows for discovery of relevant documents and communications. The use of multiple, varied search terms is essential to ensure that important documents are not missed where the terminology used by NVIDIA's employees varies across departments.

B. NVIDIA Fails Adequately to Address Structured Data and Model Data

These cases involve not only email and simple documents, but also structured data and model data from NVIDIA's AI training processes. The central issue in this litigation is NVIDIA's alleged use of Plaintiffs' copyrighted works to train its NeMo Megatron AI models. Understanding how such models used Plaintiffs' works requires discovery beyond traditional document and email production. The datasets involved are complex and encompass structured data, model parameters, training data, source code, and other technical details that reveal the inner workings of the AI models at issue.

The discovery of structured data and model data is crucial to these cases, as it will offer insight into how NVIDIA processed, transformed, and used Plaintiffs' works to train its AI models. This data likely includes model parameters that govern how the models were trained, the architecture of the models themselves, and the datasets that were fed into the training process. Examining these components allows Plaintiffs to understand not only whether their works were

unlawfully copied and used, but also how integral these works were to the creation and improvement of NVIDIA's AI models. For example, training data logs and model parameters can demonstrate how much weight was given to Plaintiffs' works during a model's training and whether the model relies heavily on these works to achieve its final outputs. Without access to this critical technical data, it will be impossible for Plaintiffs fully to establish the extent of the infringement and the degree to which their works contribute to the performance of NVIDIA's models. Plaintiffs' proposal therefore appropriately addresses these needs by specifically including the discovery of structured data and AI model-related information—elements that are crucial for understanding the scope and nature of the infringement. *See* Plaintiffs' ESI Protocol §§ 11-12; Ball Decl. ¶ 14.

NVIDIA's ESI Protocol states that the parties will meet and confer to address the production of data related to models, including training data, source code, model parameters, model architecture, version history, and other relevant information. In other words, NVIDIA opts to delay addressing structured and model data entirely, despite its centrality to the case. This failure to address key data would severely limit Plaintiffs' ability to obtain the very evidence necessary to prove their cases. *See* Ball Decl. ¶ 12 (discussing the MPO's lack of structured data provisions). NVIDIA's failure to include structured and model data in its proposal is not only a glaring omission, but also suggests an attempt to delay or even avoid producing the data at the heart of these cases. The absence of guidelines for the production of model data severely undermines the transparency of the discovery process and would leave Plaintiffs without access to the full scope of NVIDIA's AI development methods. Given the importance of this data to these cases, the lack of any meaningful framework for producing structured data effectively shields NVIDIA from having to disclose the most detrimental evidence of its infringement.

C. NVIDIA's Proposed Search Methodology is Insufficient

NVIDIA's ESI Protocol includes provisions for search term hit reports, embedded files, and a limited procedure for hyperlinked documents, but omits crucial validation procedures and fails to provide robust safeguards for ensuring transparency throughout the discovery process. The scale and complexity of discovery in these cases necessitate not only hit reports but also

comprehensive validation protocols to ensure that the search terms and methodologies are effectively capturing relevant materials. *See* Ball Decl. ¶ 12 (explaining that among the MPO's faults is its failure to address search methodologies).

The hit reports outlined in NVIDIA's proposal provide counts of documents matching search terms, unique documents, and family members requiring review, which are helpful for evaluating search term performance. *See* NVIDIA's ESI Protocol § 12. But without additional validation procedures, Plaintiffs are forced to rely solely on NVIDIA's assessment of relevance, creating the significant risk that important documents will be overlooked. Plaintiffs' proposal emphasizes the need for formal validation measures—such as quality control checks—to confirm that search terms are capturing relevant information while minimizing irrelevant results. *See* Plaintiffs' ESI Protocol § E. These checks would provide a safeguard against both underproduction and overproduction of documents, ensuring a discovery process that is both fair and proportional.

The provisions in NVIDIA's ESI Protocol for hyperlinked documents are also inadequate. NVIDIA's proposal would allow the receiving party to request specific hyperlinked documents only *after* identifying the hyperlinks and corresponding Bates numbers. *See* NVIDIA ESI Order § 17. This will place an undue burden on Plaintiffs. Given the prevalence of hyperlinked documents in corporate communications, especially through platforms like Google Drive or Microsoft Teams, NVIDIA's ESI Protocol limitation risks the exclusion of key evidence. Plaintiffs' proposal takes a more comprehensive approach by ensuring that hyperlinked documents are included as part of the standard discovery process, reducing the likelihood of missing critical evidence. *See* Plaintiffs ESI Order § 10; Ball Decl. ¶ 16.

D. NVIDIA's Proposed Approach to Mobile Data Discovery Is Not Adequate

The sharp contrast between Plaintiffs' and NVIDIA's approaches to mobile data discovery further underscores why the Court should adopt Plaintiffs' ESI Protocol. NVIDIA's treatment of mobile data is unacceptably narrow and fails to acknowledge the critical role that modern communication methods play in today's corporate environment. Key communications frequently occur outside of traditional emails and instead on mobile devices or through enterprise messaging

platforms such as WhatsApp, Signal, Slack, Workplace, or Microsoft Teams. *See* Ball Decl. ¶ 9. NVIDIA's approach, which fails adequately to account for this reality, is outdated and would artificially limit the scope of discovery by excluding essential data sources. Plaintiffs' proposed ESI protocol explicitly ensures that mobile devices, such as cell phones and tablets, are included in discovery. Plaintiff ESI Protocol § F. In a case involving allegations of systemic copyright infringement through AI training, informal and cross-departmental communications are vital for understanding NVIDIA's decision-making processes regarding the use of copyrighted materials.

E. NVIDIA's Privilege Log Protocol Does Not Ensure Sufficient Transparency

Finally, NVIDIA's treatment of privilege logs is insufficient, evasive, and fails to uphold the transparency necessary for a fair discovery process. *See* Ball Decl. ¶ 12 (explaining that among the MPO's faults is its failure to address privilege logs). NVIDIA's proposed privilege log protocol provides only the bare minimum and lacks the necessary detail to ensure that privilege claims are properly assessed. *See, e.g.*, *Prado v. Equifax Info. Servs. LLC*, No. 18-cv-02405-PJH (LB), 2019 WL 88140, at *3 (N.D. Cal. Jan. 3, 2019) ("If a party withholds material as privileged ... it must produce a privilege log that is sufficiently detailed for the opposing party to assess whether the assertion of privilege is justified."). Without robust requirements, NVIDIA could easily withhold documents by over-asserting privilege, leaving Plaintiffs without access to critical evidence. Plaintiffs' proposed protocol for privilege logs would avoid that outcome. *See* Plaintiffs' ESI Protocol § 14.

For all of the reasons stated above, Plaintiffs' proposed ESI protocol is the only proposal before the Court that is suitable for this action. The Court should therefore adopt it.

IV. The "Source Code" Designation in NVIDIA's Protective Order is Overly Broad

The primary dispute between the parties regarding the Protective Order concerns the handling of source code. While both parties agree that actual source code warrants heightened protection, they differ on how to define "Source Code." Plaintiffs argue that a clear definition of "Source Code" will ensure a shared understanding of the materials that fall within this category. In contrast, NVIDIA does not offer a specific definition but instead lists the types of documents it believes should be designated as "Highly Confidential – Source Code." *See* NVIDIA's Protective

Order § 9. NVIDIA's proposed designation, however, is overly broad and would allow documents that are not actual source code to be classified as such. Plaintiffs propose a narrower, more precise definition that would exclude materials not genuinely part of the source code, such as training data, design documents, or emails discussing source code conceptually. *See* Plaintiffs' Protective Order § 2.18.

Plaintiffs acknowledge NVIDIA's concern that certain documents, while not technically "Source Code," may still require the highest level of protection. Accordingly, Plaintiffs have included a provision requiring the parties to meet and confer about extending the "Highly Confidential – Source Code" designation to specific non-source code materials. *See id.* § 2.9. Regardless, given the stringent restrictions proposed for documents designated as "Highly Confidential – Source Code"—such as the requirement for advanced expert disclosures under Section 7.4, and the overly burdensome inspection protocols involving NVIDIA's virtual private network, which would severely limit Plaintiffs' ability to code or take notes, or worse, inspections "in a secured room without Internet access," as outlined in Section 9—Plaintiffs firmly believe the definition of "Source Code" should be narrowly tailored to include only actual source code. It should not encompass any document NVIDIA deems to have technical significance. This approach risks excessive designation, which could hinder Plaintiffs' ability to effectively examine witnesses or even review evidence related to their own works within the designated materials. In contrast, Plaintiffs' proposal strikes the appropriate balance, providing sufficient protections while making reasonable compromises on critical provisions that matter to NVIDIA.

Plaintiffs also have added language in Section 7.4 clarifying that Highly Confidential information "may be disclosed to an Expert without prior disclosure of the Expert's identity, provided the Expert is not a current officer, director, or employee of a competitor, nor is anticipated to become one." *See, e.g.*, *TVIIM, LLC v. McAfee, Inc.*, No. 13-CV-04545-VC (KAW), 2014 WL 2768641, at *2 (N.D. Cal. June 18, 2014) (noting that the Northern District of California "clearly requires that an 'expert' under the Protective Order may not be a 'past or current employee of a Party or a Party's competitor'"). Plaintiffs maintain, however, that while NVIDIA

may have a legitimate interest in vetting experts with access to their source code who are current or potential competitors, they should not have veto power over Plaintiffs' selection of experts.

Finally, Plaintiffs added a provision related to clawbacks—a provision lacking in NVIDIA's proposal—which would provide clarity on the mechanism to retrieve privileged or confidential information if any is inadvertently disclosed.

V. Conclusion

The purpose of a comprehensive ESI Protocol and Protective Order is to create standardized procedures for the collection, review, and production of relevant materials, ensuring an efficient and cost-effective discovery process. *See* Ball Decl. ¶¶ 6, 8, 17. For the reasons stated above, Plaintiffs' proposed ESI Protocol and Protective Order are the only frameworks that can ensure the discovery process in these complex copyright cases is fair, comprehensive, and proportional. NVIDIA's attempts to impose rigid, outdated limitations designed for patent cases would obstruct the very transparency and breadth of discovery that the Federal Rules of Civil Procedure mandate. The stakes in these class action copyright cases—where NVIDIA is accused of systemic infringement through AI training—are far too high to allow for an artificially narrow and overly restrictive discovery process.

Plaintiffs' ESI Protocol, which is grounded in modern electronic discovery practices, appropriately addresses the complexities of structured data, model training data, and the breadth of internal communications necessary to uncover the scope of NVIDIA's infringing conduct. It ensures a level playing field and allows the discovery of essential evidence while remaining mindful of proportionality and efficiency.

By adopting Plaintiffs' proposed ESI Protocol and Protective Order, the Court will not only promote an efficient and cost-effective process, but also safeguard the rights of Plaintiffs to obtain the evidence needed to prove their claims of copyright infringement. NVIDIA's attempt to impose undue limitations on discovery would frustrate the discovery process, conceal critical evidence, and impair Plaintiffs' ability to hold NVIDIA accountable.

Plaintiffs therefore respectfully request that the Court adopt **in full** Plaintiffs' ESI Protocol and Protective Order.

Dated: September 19, 2024

Respectfully submitted,

By: */s/ Joseph R. Saveri*

Joseph R. Saveri (State Bar No. 130064)
Christopher K.L. Young (State Bar No. 318371)
Elissa A. Buchanan (State Bar No. 249996)
Evan Creutz (State Bar No. 349728)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1505
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
cyoung@saverilawfirm.com
eabuchanan@saverilawfirm.com
ecreutz@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@buttericklaw.com

Brian D. Clark (admitted *pro hac vice*)
Laura M. Matson (admitted *pro hac vice*)
Arielle Wagner (admitted *pro hac vice*)
Eura Chang (admitted *pro hac vice*)
LOCKRIDGE GRINDAL NAUEN PLLP
100 Washington Avenue South, Suite 2200
Minneapolis, MN 55401
Telephone: (612)339-6900
Facsimile: (612)339-0981
Email: bdclark@locklaw.com
lmmatson@locklaw.com
aswagner@locklaw.com
echang@locklaw.com

Attorneys for Plaintiffs and the Proposed Class in the Nazemian *Action*

Bryan L. Clobes (*pro hac vice*)
Alexander J. Sweatman (*pro hac vice*)
Mohammed A. Rathur (*pro hac vice*)
CAFFERTY CLOBES MERIWETHER & SPRENGEL LLP
135 South LaSalle Street, Suite 3210
Chicago, IL 60603
Telephone: (312) 782-4880
bclobes@caffertyclobes.com
asweatman@caffertyclobes.com
mrathur@caffertyclobes.com

Amy E. Keller (*pro hac vice*)
Nada Djordjevic (*pro hac vice*)
James A. Ulwick (*pro hac vice*)
DiCELLO LEVITT LLP
Ten North Dearborn Street, Sixth Floor
Chicago, Illinois 60602
Tel. (312) 214-7900
akeller@dicellolevitt.com
ndjordjevic@dicellolevitt.com
julwick@dicellolevitt.com

David A. Straite (*pro hac vice*)
DiCELLO LEVITT LLP
485 Lexington Avenue, Suite 1001
New York, NY 10017
Tel. (646) 933-1000
dstraite@dicellolevitt.com

Attorneys for Plaintiffs and the Proposed Class in the Dubus *Action*